

Defence against adversarial attacks on IoT detection systems using deep belief network

Sharipuddin¹, Eko Arip Winanto²

¹Department of Informatics, Faculty of Computer Sciences, Universitas Dinamika Bangsa, Jambi, Indonesia

²Department of Computer Engineering, Universitas Dinamika Bangsa, Jambi, Indonesia

Article Info

Article history:

Received Jan 29, 2024

Revised Apr 1, 2024

Accepted Apr 6, 2024

Keywords:

Adversarial attack

Deep belief network

FGSM

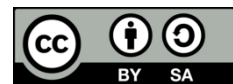
Internet of things

Intrusion detection system

ABSTRACT

An Adversarial attack is a technique used to deceive machine learning models to make incorrect predictions by providing slightly modified inputs from the original. Intrusion detection system (IDS) is a crucial tool in computer network security for the detection of adversarial attacks. Deep learning is a trending method in both research and industry, and this study proposes the use of a deep belief network (DBN). DBN can recognize data with small differences, but is also vulnerable to adversarial attacks. Therefore, this research suggests an internet of things-intrusion detection system (IoT-IDS) architecture using a DBN that can counter adversarial attacks. The chosen adversarial attack for this study is the fast gradient sign method (FGSM) used to evaluate the IoT IDS using the DBN model. Testing was conducted in two scenarios: first, the model was trained without adversarial attacks; second, the model was trained with adversarial attacks. The test results indicate that the DBN model struggles to detect FGSM attacks, achieving an accuracy of only 46% when it is not trained with adversarial attacks. However, after training with the FGSM dataset, the DBN model successfully detected adversarial attacks with an accuracy of 97%.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Sharipuddin

Department of Informatics, Faculty of Computer Sciences, Universitas Dinamika Bangsa

36125 Kota Baru, Jambi, Indonesia

Email: sharipuddin@unama.ac.id

1. INTRODUCTION

The success of deep learning (DL) in detecting subtle changes, such as small pixel alterations in images, demonstrates its reliability during the training process [1], [2]. Research indicates that the application of DL is not limited to big data, but can also be implemented in network traffic classification and intrusion detection systems (IDS) [3]. Several prior studies [4]-[6] have utilized deep learning; for instance, in [7], a deep belief network (DBN)-based approach was proposed for internet of things (IoT) network attack detection, achieving an accuracy of 99.0%. Similarly, [8] proposed a DBN method for IoT networks. Additionally, [9] introduced a hybrid model to enhance the DBN performance with an autoencoder, revealing an improved detection performance compared to a standalone DBN.

A recent study indicated that deep learning systems are susceptible to carefully crafted adversarial attacks with subtle and imperceptible effects. This raises concerns regarding the reliability of systems that employ DL [10]. A potential solution is the use of deep-learning-based intrusion detection systems. However, these detectors are vulnerable to adversarial attacks, in which features are manipulated to allow the manipulated network features to evade detection. Adversarial training is a technique employed to develop robust intrusion detectors using a min-max formulation. In adversarial training, attack samples are manipulated using various adversarial attacks to produce poisoned attack samples. These poisoned attack

samples were incorporated into the model training to enhance their resilience against evasion attacks (i.e., attacks during testing).

Hence, a challenge that needs improvement is enhancing the performance of the DBN for IDS in IoT networks against adversarial attacks. Chen *et al.* [10], some neural network-based malware IDS were developed, with nine IDS trained with specific adversarial attacks and one without adversarial attack training data. The attacks consisted of fast gradient sign method (FGSM), DeepFool-guided sign method with k steps (dFGSMk), Deepfool and grosse adversarial. The experimental results indicated that this novel approach significantly improved the resilience of the malware detector, achieving the lowest evasion rate of 12%.

In a subsequent study [11], the identification of radio frequency (RF) fingerprints based on DL has demonstrated the resilience of the corresponding DL model against adversarial attacks and defenses. This study explores the effects of four recent adversarial attacks on DL-based RF fingerprint identification and provides a graphical analysis of RF feature disruption following adversarial attacks. The results, compared with several common defense methods, validate the capability to enhance accuracy to 96.00%, even when the adversarial attack interference is 5.00%, including physical attacks.

Furthermore, a recent study [12] indicated that medical deep learning systems can be affected by carefully crafted adversarial examples with subtle imperceptible disturbances. This raises safety concerns regarding the implementation of these systems in clinical settings. The findings revealed that medical deep neural network (DNN) models are more susceptible to adversarial attacks than models for natural images from two different perspectives.

Therefore, there is a need for an evaluation to enhance resilience against and counter adversarial attacks in IoT detection systems by using the DBN method. This work focuses on defense mechanisms and strategies to counter adversarial attacks on IoT-IDS employing a deep belief network. In addition, this research makes several contributions: i) constructing a dataset in an IoT network with embedded adversarial attacks; ii) proposing a detection system against adversarial attacks on IoT networks using DBN; and iii) identifying the effects of adversarial attacks on the performance of an IDS using DBN. This paper is organized into four main sections. Section 1 provides the introduction. Section 2 offers a concise overview of the experimental dataset and setup of the study. Section 3 provides a more detailed description of the experiment and findings of the study. Finally, section 4 summarizes the conclusions and proposes potential directions for future research.

2. METHOD

This study focuses on an approach to address adversarial attacks using DL. In this section, the steps for completing this study are outlined. It encompasses datasets, experimental configurations, feature selection techniques, adversarial attack datasets, classification algorithms, and experimental tools.

2.1. Dataset

In this study, the CICIOT2022 dataset constructed by the University of New Brunswick, Canada, was utilized. The CIC IoT 2022 dataset is specifically designed for research in the context of the IoT. This dataset provides information related to the security and performance of the IoT devices. The data within the dataset include various variables such as IP addresses, network protocols, timestamps, and other relevant attributes for security analysis, comprising 47 features. In addition, the dataset may encompass scenarios and recorded attacks to facilitate in-depth security research. By leveraging the CIC IoT 2022 dataset, this study aims to develop deep-learning models to counter adversarial attacks. Table 1 lists the features of this dataset. This dataset encompasses 34 types of attacks, including DDoS, BenignTraffic, Mirai, Recon, and SqlInjection [13].

Table 1. Dataset

No	Dataset	Numbers of record
1	141	223,999
2	142	260,694
3	143	239,202
4	144	244,859
5	145	442,721
6	146	227,148
7	147	237,858
8	148	235,009
9	149	445,891
10	150	232,166

2.2. Experiment setup

This research places significant emphasis on the experimental phase concerning IDSs that utilize DL to mitigate adversarial attacks. These experiments entail a thorough investigation of various aspects, including the creation of the adversarial attack dataset, the application of feature selection techniques, the distribution of datasets for training and testing, and the design and configuration of the DBN model and its associated variables. In general, the experimental setup comprises three stages, as illustrated in Figure 1, which can be delineated as follows:

- Constructing the adversarial attack dataset, wherein the FGSM is employed to build the dataset and information gain is utilized for feature selection.
- Subsequently, regular and adversarial attack datasets were classified using the DBN. The results were analyzed by considering parameters such as the validation and testing accuracy.
- Finally, a comparison and analysis were conducted on the accuracy of the training, validation, testing, and adversarial attacks for each dataset type.

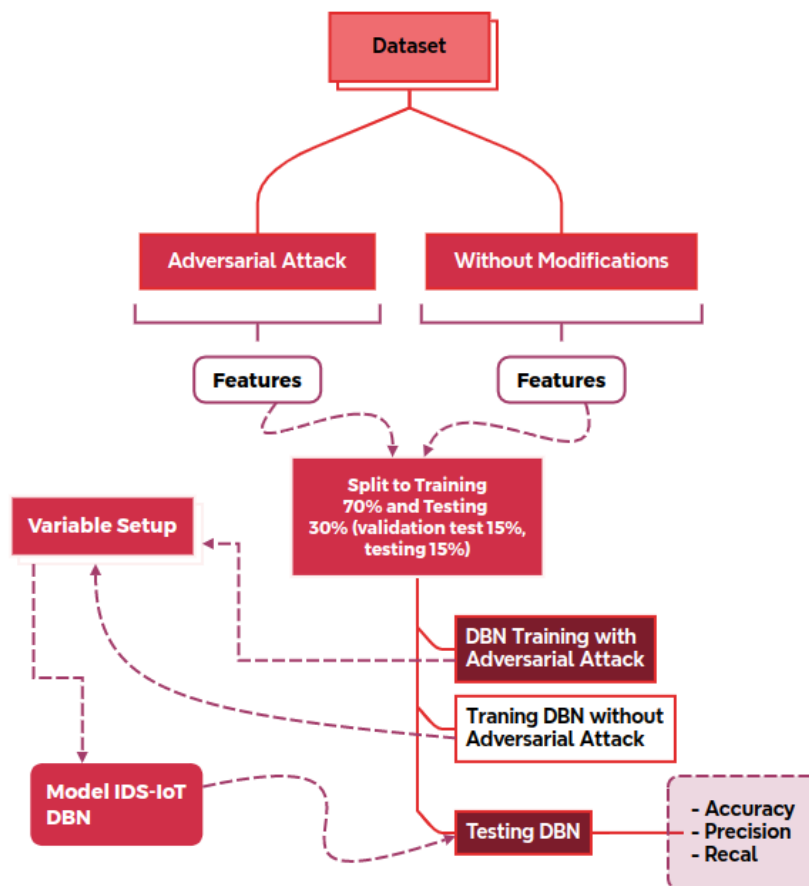


Figure 1. Experiment design

2.3. Feature selection using information gain

The widely adopted feature selection technique of information gain operates as an information gain filter [14]. This method aims to diminish the noise caused by irrelevant features by employing simple attribute ranking, subsequently pinpointing the features that carry the majority of information within a given class [15]-[18]. Evaluating the entropy of the features serves as a criterion for identifying superior features. Entropy, as a measure of uncertainty, successfully portrays feature distributions [19]. The selection of information gain as the feature selection method for this study is justified by its filter-based nature, offering a more robust set of selected features that effectively combats overfitting [20]. Therefore, employing feature selection techniques that yield significant, pertinent features with reduced computational complexity contributes to a reduction in execution time for the classification algorithms utilized in attack detection on IoT networks.

2.4. Constructing an adversarial attack dataset

When constructing adversarial samples, it is crucial to consider factors such as the complexity of the input data, the model employed, and the necessary parameters to generate effective samples. This consideration plays a significant role in enhancing the security of IoT networks and minimizing the likelihood of adversarial attacks [21]. The adversarial attack samples in this research were constructed using the FGSM. FGSM is an adversarial attack method for machine learning models that is used to evaluate model security. In this context, 'eps=0.1' is employed, where epsilon (ϵ) is set to 0.1, denotes the maximum allowable change in each input feature when the attack is executed. Technically, FGM involves calculating the gradient of the loss function with respect to the input, and then adjusting the input in the direction of that gradient. Epsilon acts as a scalar that controls the magnitude of the changes allowed for each feature. By setting epsilon to 0.1, this attack is constrained to make relatively small changes to each input element.

2.5. Deep belief network

DBN, a type of DNN, is composed of layers of restricted Boltzmann machines (RBM) stacked together. Proposed by [22], [23], this generative model can be applied to perform unsupervised learning tasks for feature dimension reduction and can also be utilized for supervised learning tasks to construct classification or regression models. Training a DBN involves two steps: layer-by-layer training and fine-tuning [24]. Layer-by-layer training entails unsupervised training of each RBM, while fine-tuning involves employing an error back-propagation algorithm to adjust DBN parameters after completing the unsupervised training [25], [26]. The DBN model representing the combined distribution between the observed vector x and l hidden layers h_k is depicted in (1).

$$P(x, h^1, \dots, h^l) = (\prod_{k=0}^{l-2} P(h^k | h^{k+1})) P(h^{l-1}, h^1) \quad (1)$$

Where $x=h^0$, $P(h^k | h^{k+1})$ is the conditional distribution for visible units conditioned on hidden RBM units at level k , and $P(h^{l-1}, h^1)$ is the combined visible-hidden distribution in top-level RBMs.

2.6. Analysis tools

The simulations were executed on a computer equipped with an Intel Core i7 processor operating at 2.60 GHz and 12 GB RAM, running on the Ubuntu 20.04.3 LTS operating system. For the purpose of analysis using scikit-learn and TensorFlow were utilized for deep learning. Additionally, the hard library was employed for deep learning.

3. RESULTS AND DISCUSSION

This extensive analysis delineates the creation of the adversarial attack dataset, the approach used for feature selection and normalization, the deployment of the IDS-IoT DBN model to counter adversarial attacks, and a comprehensive examination of the experimental findings, incorporating a detailed discussion of the results. This section elaborates on the experimental outcomes. Furthermore, it conducts an analysis of the experimental results and contrasts them with the findings of prior studies.

3.1. Result of information gain

In this section, we propose the use of feature selection to reduce the number of features considered initially. The objective is to obtain the most important features in the IoT IDS detection process using the DBN. Table 2 presents the results of feature selection using information gain (IG), which were sorted along with the weight values derived from IG calculations. In this study, the top 20 features were selected from the initial 47 features based on IG calculations. These features are then utilized for the training and testing processes of the IDS detection system. The outcome of the subsequent feature selection process involved transforming the values of each feature in the dataset to a smaller scale. The purpose of this step is to reduce the computational burden of the detection system by using a DBN. At this stage, the values of the dataset were rescaled to a range from -0 to 1.

3.3. Results of adversarial attack dataset

The FGSM is an adversarial attack method on machine learning models that is used to evaluate model security. In this context, 'eps=0.1' is employed, where epsilon (ϵ), set to 0.1, denotes the maximum allowable change in each input feature when the attack is executed. Technically, FGM involves calculating the gradient of the loss function with respect to the input and then adjusting the input in the direction of that gradient. Epsilon acts as a scalar controlling the magnitude of changes allowed in each feature. By setting epsilon to 0.1, this attack is constrained to make relatively small changes to each input element. Configuring

epsilon in this manner has significant implications in the context of model security because it enables measuring the resilience of a model against attacks with limited changes to the input. Attacks with small epsilon values provide insights into the model's resistance to attacks intended to induce minor yet impactful changes in model predictions. Table 3 illustrates an example of the dataset value change as input, with an epsilon value of 0.1.

Table 2. Feature selection using information gain

Features ID																																											
39	38	1	41	36	33	34	35	2	26	15	0	4	5	18	8	27	17	30	43	37	42	16	14	11	44	7	10	9	3	45	40	20	19	12	32	31	23	28	21	13	24	25	6
29 22																																											
Weight of features																																											
2.5979944476613723, 1.328664984044619, 1.3203696712811852, 1.305611534185723, 1.3017253080487134, 1.2942106754577085, 1.2859045680429388, 1.2571670518394447, 1.1582000821827094, 0.649369983292369, 0.6466184950292497, 0.6386970554784903, 0.622151957880436, 0.6221216072432929, 0.5064510320445264, 0.49804177837213626, 0.49312590226614583, 0.4602672807341217, 0.4384311714442015, 0.3720382902693218, 0.37120629824615303, 0.37028579893864855, 0.35001815704680794, 0.3308688289409405, 0.3255047709636183, 0.29364815604746264, 0.2933656156744644, 0.2869057737353793,																																											

Table 3. Adversarial attack dataset

Original dataset values		Adversarial attack dataset values	
-0.00825567	-0.29309016	0.09174433	-0.19309017
-0.16642086	-0.31401501	-0.06642087	-0.214015
-0.29125409	-0.28425318	-0.19125411	-0.18425319
-0.26689492	-0.24201057	-0.16689493	-0.34201056
-0.34193691	0.85445358	-0.4419369	0.9544536

3.4. Result of IDS-IoT using DBN

To analyze the performance of the IDS-IoT model with a DBN in addressing adversarial attacks, measurements such as accuracy in the training, validation, testing, and adversarial attack testing stages were employed. In the experiments, each IG feature subset was classified by DBN into 20 features. In this study, two steps were conducted to utilize a DBN for intrusion detection in IoT networks. The first step involved a learning process using both regular and adversarial attack datasets. The variable settings for the DBN model are listed in Table 4. The outcomes of the DBN learning process include the weight and bias values that are used in the detection or prediction process. Subsequently, the prediction process utilizes the DBN to detect attacks on IoT networks.

In the testing process, multiple testing scenarios were conducted with 10 datasets (csv), and the accuracy parameters were observed at each stage. Table 5 displays the testing results of the DBN model, starting from the training, validation, testing, and adversarial attack datasets. The testing outcomes revealed a high accuracy at each stage, except for the adversarial attack dataset. From these test results, it can be concluded that the IDS with adversarial attacks significantly affects the accuracy of the DBN model. The accuracy of adversarial attack testing only reached 46% compared to the testing datasets without adversarial attacks.

These results indicate that the DBN model is vulnerable to adversarial attacks applied to the datasets. The fast gradient method (FGM) is used in the context of computer security and machine learning to assess and enhance model security. The parameter $\epsilon=0.1$ in the context of FGM represents the epsilon value used in crafting attacks on the model. Epsilon signifies the maximum allowable change in each input feature when an attack is executed. In the context of attacks on machine learning models, such as the FGM method, the epsilon value is used to control the magnitude of the changes permitted in the input features. For example, if epsilon is set to 0.1, it means that the attack can only induce changes up to 0.1 on each input feature, maintaining these changes within the specified bounds. Using a small epsilon value, attacks can be modeled as slight disturbances to the input, aiding in understanding the resilience of the model against various levels of adversarial attacks. Adjusting the epsilon value can be used to test the resistance of the model against different attack magnitudes.

This study proposes the mixing of datasets to address adversarial attacks on IoT IDS. Addressing adversarial attacks by combining regular datasets with datasets generated from adversarial attacks is a strategy that can be applied to enhance the model's resilience against attacks. In this study, the blending of datasets generated through the FGM and regular datasets aims to introduce diversity into the training data, enabling the model to learn more general patterns and become less dependent on the specific characteristics of adversarial attacks.

Table 4. Variable DBN

Variable name	Description
Number of layers	4(1 input, 1 hidden, 1 output)
Node	100 node, 64 node, 34 node
Input dimension	Relu, Relu, Relu, sigmoid
Output dimension	20
Epoch	50
Batch size	32

Table 5. The results of IDS-DBN without adversarial attacks

Dataset	Accuracy of training	Accuracy of validation	Accuracy of testing	Accuracy of adversarial attack
141	0.9988	0.9987	0.9970	0.23
142	0.9991	0.9990	0.9780	0.14
143	0.9989	0.9991	0.9957	0.46
144	0.9990	0.9990	0.9808	0.36
145	0.9990	0.9991	0.9989	0.05
146	0.9990	0.9990	0.9950	0.21
147	0.9990	0.9989	0.9914	0.36
148	0.9988	0.9983	0.9987	0.07
149	0.9990	0.9990	0.9980	0.08
150	0.9989	0.9990	0.9984	0.12

The process begins with the creation of two sets of data: a regular dataset containing original samples, and a dataset resulting from adversarial attacks using FGM. Subsequently, these two datasets were mixed or merged into a single dataset encompassing variations of both types of data. Thus, it is anticipated that the DBN model can be trained to recognize patterns from both regular and adversarial attack data simultaneously.

The goal of dataset mixing is to help the model become more tolerant to small variations in the input and minimize the impact of adversarial attacks. This is because the model learns not only from patterns present in regular datasets, but also recognizes variations introduced by adversarial attacks. Consequently, the model can become more robust and provide more stable predictions for various types of inputs.

Table 6 displays the test results after training the model using both the datasets. The results show that the model can detect small changes from attacks occurring in an IoT-detection system. This is evident in the testing results, where the accuracy of adversarial attack detection improved compared to previous tests without training with adversarial attack datasets. Although the accuracy results are not yet optimal, there is a need for optimization to improve the detection outcomes of adversarial attacks.

Table 6. The results of IDS-DBN without adversarial attacks

Dataset	Accuracy of training	Accuracy of validation	Accuracy of testing	Accuracy of adversarial attack
141	0.9987	0.9988	0.9981	0.9648
142	0.9990	0.9989	0.9795	0.9678
143	0.9990	0.9989	0.9984	0.9659
144	0.9990	0.9990	0.9823	0.9722
145	0.9998	0.9998	0.9961	0.9674
146	0.9990	0.9989	0.9990	0.9585
147	0.9990	0.9988	0.9823	0.9664
148	0.9990	0.9989	0.9990	0.9670
149	0.9989	0.9988	0.9984	0.9640
150	0.9989	0.9988	0.9943	0.9536

3.5. Analysis

The implementation of adversarial attacks on IDS-IoT DBN can be observed in previous tables. The next step is to analyze the accuracy results of the training and testing processes, as well as the impact of implementing adversarial attacks on the IDS-DBN in the IoT network. Figure 2 shows a graph depicting the training accuracy over 50 epochs. The graph shows that the training process approached a value close to 1, indicating the optimal performance of the training results. Additionally, the validation accuracy is represented to ensure that the training results do not suffer from overfitting.

Figure 3 displays the training loss during the training phase of the DBN model in the IoT-IDS detection system. From the graph, it can be inferred that the loss values decreased as the number of epochs increased. This indicates that the training process was progressing well, suggesting that the error diminished.

Following the completion of the training process and obtaining the DBN model for the IoT detection system, the subsequent step involved testing. This test was conducted to detect attacks occurring in an IoT network. Figure 4 shows a heatmap matrix illustrating the results of testing the IoT-IDS detection system using a DBN. The diagonal colors in the image represent the values of the tests that were detected correctly.

The final conclusion from this testing indicates that the adversarial attack, specifically FGSM, has the potential to significantly impact the performance of the IDS-DBN model. This was attributed to the failure of the DBN model to recognize small changes resulting from adversarial attacks. The approach employed in this study to address adversarial attacks involves blending regular datasets with adversarial attack datasets. The results of this approach successfully identify adversarial attacks occurring in IoT networks using the DBN model.

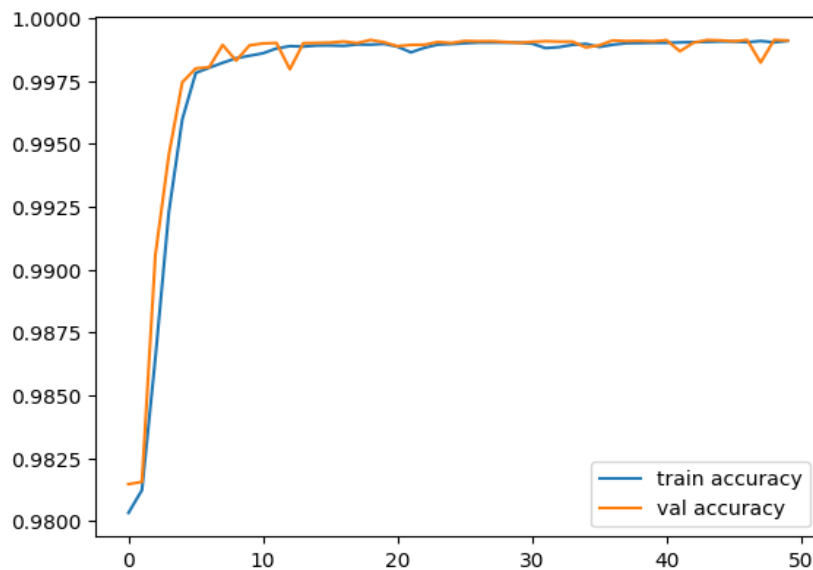


Figure 2. Result of training accuracy

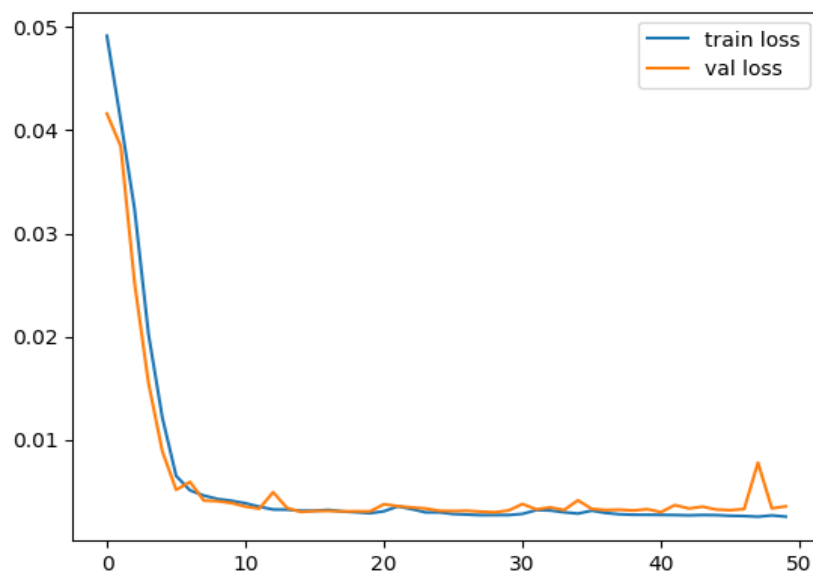


Figure 3. Result of training loss

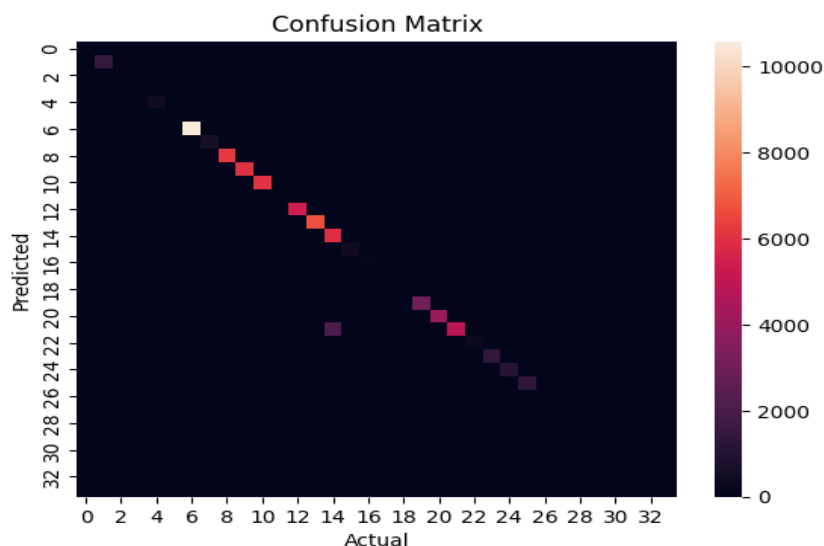


Figure 4. Matrix test results

4. CONCLUSION

DL has been implemented in various fields including network security. However, challenges can affect the performance of DL, such as adversarial attacks. This study constructed an adversarial attack dataset using the FGSM with $\epsilon=0.1$. The results indicate that the generated dataset can disrupt the performance of the IDS detection system, achieving an accuracy of only 46 %. The proposed IoT detection system model employs the DBN method to counter adversarial attacks. This approach involves mixing IoT datasets with datasets modified using FGSM. Testing the model reveals that the DBN can detect small changes from adversarial attacks, achieving an accuracy of 97% after training with both datasets. In conclusion, adversarial attacks significantly affect the performance of a detection system that uses the DBN model. The DBN model struggles to detect the small changes caused by an attack. Future research challenges include proposing alternative approaches to address adversarial attacks on deep-learning models in IoT networks. Additionally, it is crucial to explore other types of adversarial attack testing in the designed IDS system.

ACKNOWLEDGEMENTS

This research is supported by Dinamika Bangsa University through the human resources development program with contract number 015/MOU/LPPM/UNAMA/VIII/2023. The objective of this program is to enhance the quality of human resources at Dinamika Bangsa University through research and development.

REFERENCES

- [1] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for internet of things," *Future Generation Computer Systems*, vol. 82, pp. 761–768, May 2018, doi: 10.1016/j.future.2017.08.043.
- [2] A. Alrawais, A. Alhothaily, C. Hu, and X. Cheng, "Fog computing for the internet of things: security and privacy issues," *IEEE Internet Computing*, vol. 21, no. 2, pp. 34–42, 2017.
- [3] S. Sharipuddin *et al.*, "Intrusion detection with deep learning on internet of things heterogeneous network," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, pp. 735–742, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp735-742.
- [4] M.-J. Kang and J.-W. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," *PLOS ONE*, vol. 11, no. 6, p. e0155781, Jun. 2016, doi: 10.1371/journal.pone.0155781.
- [5] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, 2016, pp. 21–26, doi: 10.4108/eai.3-12-2015.2262516.
- [6] Y. Li, R. Ma, and R. Jiao, "A hybrid malicious code detection method based on deep learning," *International Journal of Security and Its Applications*, vol. 9, no. 5, pp. 205–216, May 2015, doi: 10.14257/ijisa.2015.9.5.21.
- [7] A. Thakkar and R. Lohiya, "A review on machine learning and deep learning perspectives of IDS for IoT: recent updates, security issues, and challenges," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 3211–3243, 2021, doi: 10.1007/s11831-020-09496-0.
- [8] N. Balakrishnan, A. Rajendran, D. Pelusi, and V. Ponnusamy, "Deep belief network enhanced intrusion detection system to prevent security breach in the internet of things," *Internet of Things*, vol. 14, p. 100112, Jun. 2021, doi: 10.1016/j.iot.2019.100112.

- [9] Sharipuddin *et al.*, "Enhanced deep learning intrusion detection in IoT heterogeneous network with feature extraction," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 9, no. 3, pp. 747–755, 2021, doi: 10.52549/V9I3.3134.
- [10] Y. Chen, C. Kreitzer, and D. Song, "Boundary attack: a simple and effective method to fool deep neural networks," in *2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4929–4938.
- [11] W. Brendel and M. Bethge, "Approximating CNNs with bag-of-local-features models works surprisingly well on ImageNet," in *7th International Conference on Learning Representations, ICLR 2019*, 2019.
- [12] S.-M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, "Universal adversarial perturbations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1765–1773.
- [13] S. Dadkhah, H. Mahdikhani, P. K. Danso, A. Zohourian, K. A. Truong, and A. A. Ghorbani, "Towards the development of a realistic multidimensional IoT profiling dataset," in *2022 19th Annual International Conference on Privacy, Security & Trust (PST)*, Aug. 2022, pp. 1–11, doi: 10.1109/PST55820.2022.9851966.
- [14] M. H. Aghdam and P. Kabiri, "Feature selection for intrusion detection system using ant colony optimization," *International Journal of Network Security*, vol. 18, no. 3, pp. 420–432, 2016.
- [15] S. Sharipuddin, E. A. Winanto, Z. Z. Mohtar, K. Kurniabudi, I. S. Wijaya, and D. Sandra, "Improvement detection system on complex network using hybrid deep belief network and selection features," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 1, pp. 470–479, Jul. 2023, doi: 10.11591/ijeecs.v31.i1.pp470-479.
- [16] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.
- [17] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE Transactions on Computers*, vol. 65, no. 10, pp. 2986–2998, Oct. 2016, doi: 10.1109/TC.2016.2519914.
- [18] T. Ait Tchakoucht and M. Ezziyyani, "Building a fast intrusion detection system for high-speed-networks: probe and DoS attacks detection," *Procedia Computer Science*, vol. 127, pp. 521–530, 2018, doi: 10.1016/j.procs.2018.01.151.
- [19] T. A. Alhaj, M. M. Siraj, A. Zainal, H. T. Elshoush, and F. Elhaj, "Feature selection using information gain for improved structural-based alert correlation," *PLOS ONE*, vol. 11, no. 11, p. e0166017, Nov. 2016, doi: 10.1371/journal.pone.0166017.
- [20] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *Journal of Computational Science*, vol. 25, pp. 152–160, Mar. 2018, doi: 10.1016/j.jocs.2017.03.006.
- [21] S. M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "DeepFool: A simple and accurate method to fool deep neural networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 2574–2582, 2016, doi: 10.1109/CVPR.2016.282.
- [22] G. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009, doi: 10.4249/scholarpedia.5947.
- [23] S. Abirami and P. Chitra, "Energy-efficient edge based real-time healthcare support system," in *Advances in Computers*, vol. 117, no. 1, 2020, pp. 339–368.
- [24] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, p. 11, 2014.
- [25] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, p. 28, 2017.
- [26] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *2017 IEEE Symposium on Security and Privacy (SP)*, May 2017, pp. 39–57, doi: 10.1109/SP.2017.49.

BIOGRAPHIES OF AUTHORS



Dr. Sharipuddin     received a Doctor of Engineering from Universitas Sriwijaya. He is currently a Senior Lecturer at the Faculty of Computer Science, Universitas Dinamika Bangsa, Indonesia. His research interests include information technology and information security. He can be contacted at email: sharipuddin@unama.ac.id.



Eko Arip Winanto     received the B.Sc. degree in computer science from the University of Sriwijaya, Indonesia, the M.Phil. degree in computer science from the Universiti Teknologi Malaysia, Malaysia. He is currently a lecturer at the Faculty of Computer Science, Universitas Dinamika Bangsa, Indonesia. His research interests include IoT, machine learning, blockchain, and network security. He can be contacted at email: ekoaripwinanto@unama.ac.id.